质性研究中的资料饱和及其判定*

杨莉萍 亓立东 张博
(南京师范大学心理学院，南京 210024)

摘要 在质性研究中，资料饱和通常被用以评估研究资料的充足性。但在研究实践中，资料饱和却存在概念模糊及操作性问题。作为某项质性研究所抽取的样本量已满足研究需要的标志，按照在研究过程中判定时点的先后顺序，资料饱和可划分为数据饱和、编码或主题饱和、意义饱和及理论饱和四种主要形式。四种形式的资料饱和各有其特定的内涵、评估方式和判定标准。研究认为，达到资料饱和的样本量标准不应统一设定，其检验需嵌入具体的研究过程；资料饱和存在逻辑上的不确定性，适度追加抽样有助于进一步确认；资料饱和作为考察研究质量的一项重要指标，并不适用于所有质性研究。

关键词 质性研究, 资料饱和, 判别方式, 判定标准

1 引言

自 2018 年 1 月中国心理学会心理学质性研究专业委员会正式设立以来，采用质性研究及混合研究方法的论文大量增加。在质性研究中，与量化研究一样，样本充足是研究效度的基本保证(Curtis et al., 2000)。样本量不足会影响研究质量，而样本量过大又会浪费研究资源。资料饱和则是质性研究中样本量适当的标志。研究过早终止数据搜集，达不到资料饱和，会导致研究结果缺乏必要的洞见性，沦为对原始数据的简单呈现(苏达比，2006/2016)。达到资料饱和意味着在当前所搜集和分析的资料基础上，进一步的数据搜集不会帮助研究者对故事或理论产生更深入的理解，因此没有必要继续搜集和分析数据(Corbin & Strauss, 2014)。这里需要特别说明的是，与量化研究不同，质性研究中的资料搜集和资料分析并不是两个截然划分的研究阶段，而是呈现循环往复的特点。质性研究中的资料饱和往往是在对前期所搜集的资料进行分析的基础上所做出的判断。而一旦资料搜集达到饱和，就意味着资料分析也同时告一段落。

以往多以两种方式判定质性研究的样本量，或依据经验法则，或通过资料饱和检验(Marshall et al., 2013)。基于经验法则，已有研究总结出了不同类型质性研究适合的样本量，

收稿日期: 2021-07-01
* 江苏省社会科学基金项目(19SHB007)。
通讯作者：杨莉萍, E-mail: lpy2908@163.com

但是这类文献大多缺乏对基本原理的清晰描述，其所建议的样本量标准往往差异很大。例如，在研究者给出的建议中，扎根理论研究的访谈样本量范围在 5~35 之间，而案例研究(case study)的样本量范围则是在 4~30 之间(Sim et al., 2018)。更有研究者认为，由于质性研究的特征以及不同研究之间的差异性，为质性研究设置统一的样本量标准缺乏理论与现实依据(Guest et al., 2006; Marshall et al., 2013)。

将资料饱和作为质性研究中停止数据搜集和分析的方法论原则被广泛接纳与使用(Saunders et al., 2018)。资料饱和甚至被称为确定质性研究样本量的“黄金准则”与“行动指南”(Guest et al., 2006)，以及质性研究质量的“保证书”(Morse, 2015)。如有研究者提出，在质性研究中，评估抽样充分性的最常见标准是资料饱和(Fusch & Ness, 2015; Morse, 2015)；同时，资料饱和也是评判质性研究学术论文质量的重要准则(Hennink et al., 2017)，恰当使用资料饱和标准可以提高质性研究的质量，如严谨性、信效度等(Hennink et al., 2019; O’reilly & Parker, 2013; Sim et al., 2018)。美国心理协会出版物和通讯委员会关于质性研究的期刊文章报告标准工作组在制定心理学质性研究期刊论文报告标准时也指出，研究者应该说明确定研究样本量的依据，并建议采用资料饱和进行检验(Levitt et al., 2018)。

尽管资料饱和的概念与作用受到质性研究者的广泛认同并被不断提及，但在实际操作过程中，这一概念和判定标准却始终含糊不清(Hennink et al., 2017)，长期存在一系列概念性与操作化问题(Saunders et al., 2018)。一方面，资料饱和的概念源于扎根理论，最初以理论饱和(theoretical saturation)的形式出现(Glaser & Strauss, 1967)。但随着资料饱和作为质性研究样本量判断标准的推广，又发展出多种指标，例如数据饱和(data saturation)、编码饱和(code saturation)、主题饱和(thematic saturation)、意义饱和(meaning saturation)等。由于饱和的形式和标准多样，不同形式和标准之间既有差异又相互联系，关系错综复杂，以至于很少有研究者能够系统深入地理解和掌握资料饱和的概念，在使用过程中常常出现混乱(O’Reilly & Parker, 2013)。另一方面，在众多研究报告中，研究者都声称该研究以资料饱和作为样本量的判断标准，且已达到资料饱和，但是却极少说明达到的是哪种资料饱和、其判断方式及合理性(Constantinou et al., 2017; Morse, 2015)。有研究者因而提出，应增加对质性研究中资料饱和报告的透明度，在作者宣称达到了资料饱和的同时，需要提供更加细致的描述，明确资料饱和的形式、判定方式与依据，以便于读者评估和进一步验证，这也是对质性研究规范意识的基本要求(Fusch & Ness, 2015; Hennink et al., 2017; Kerr et al., 2010; Malterud et al., 2016)。但由于以往研究尚缺乏对资料饱和判定及报告的操作化描述与实践性指导(Carlsen &

Glenton, 2011; Hennink et al., 2019)，要求研究者做到这些并不现实。考虑到资料饱和对于质性研究的重要性及操作现状，有必要对各种不同形式的资料饱和的定义、判定标准与报告方式加以探究与说明，以帮助研究者更好地理解与使用资料饱和的概念，使对质性研究样本量的判断有据可循，从而提高心理学中质性研究的规范性和研究质量。

本研究基于对资料饱和相关文献的研究及作者团队长期开展质性研究的实践经验，试图明确质性研究中资料饱和的概念，厘清不同形式的资料饱和，通过比较、明确不同形式的资料饱和之间的关系，探索质性研究资料饱和的判定及报告方式，以期为质性研究中资料饱和的判断提供参考。

2 质性研究中四种形式的资料饱和

资料饱和是一项质性研究所抽取的样本量已满足研究需要的标志。最早提出资料饱和问题的是扎根理论，所采用的判断标准是理论饱和(Glaser & Strauss, 1967)。但随着质性研究的发展，资料饱和的概念不再局限于扎根理论提出的理论饱和，而是进一步发展出了包括数据饱和、编码或主题饱和、意义饱和等在内的多种形式的资料饱和概念。不同形式的资料饱和其意涵、指向、所使用的分析手段及判定标准各不相同(Saunders et al., 2018)。

2.1 理论饱和 (theoretical saturation)

在质性研究中，扎根理论最早提出了研究资料的饱和问题。扎根理论是在系统搜集资料的基础上，寻找反映社会现象的核心概念，通过在这些概念之间建立联系而形成理论的一种研究方法(陈向明, 2000)。扎根理论研究的主旨在于建构理论，作为研究结果，所建构的理论在其内部及其与外部相关理论之间需要获得一致性和协调性。基于此，扎根理论的创始人Glaser 和 Strauss(1967)最早提出了理论饱和的概念，并将其解释为“如果继续追加抽样，不会再有新的类属或相关主题出现”。

扎根理论中的理论饱和概念是指在数据搜集过程中已发展不出新的主题(issues)或理解(insights)，所有相关概念类属(conceptual categories)的属性及其之间的关系都已被探讨、穷尽和确定，继续获取数据资料已不能揭示新属性，也不能获得对于新生理论(emerging theory)更深入的理解，新的理论已全面、可信(Hennink et al., 2017; Morse, 2015)。Hennink 等(2019)指出，理论饱和依赖样本的充分性，要求研究者尽可能发现充足的、丰富的、合乎逻辑的、有意义的数据，以此支持新生理论。Saunders 等(2018)则认为，当所有表征理论的概念都充分反映在了数据中时，便达到了理论饱和。

扎根理论对理论饱和的判断植根于质性研究循环往返、螺旋式演进的过程，与扎根理论对理论抽样(theoretical sampling)的要求紧密相关(Saunders et al., 2018)。在扎根理论研究过程中，资料搜集到一定程度，并且经历了编码、比较、归类、再比较等资料分析过程之后，初步勾勒出了新生理论的轮廓，以此指导研究者进一步实施理论抽样，然后基于新搜集的资料，检验、改进和完善新生理论，如此循环演进，逐步去除新生理论中的薄弱环节。这是一个将新生理论返回原始资料进行比较、验证，不断优化，使之完善的持续性过程(陈向明, 2000)。直到新生理论基本可以解释所有的资料，研究者已不能从资料中继续发掘新的表征理论的相关概念，而只是重复表征已有概念或属性，这时便视为达到了理论饱和，可停止抽样。

有研究者指出，在扎根理论研究过程中，在理论抽样的驱动下，概念类属或新生理论循环演进，不断完善，当概念类属之间的关系及其含义逐渐清晰时，就达到了理论饱和 (Morse, 2015)。由此可以看出，理论饱和是基于理论完整性进行的饱和度判断，判断的时点在资料搜集和分析的后期阶段，具有较高的理论概括性水平(Saunders et al., 2018)。

2.2 数据饱和(data saturation)

数据饱和是指随着资料搜集的进程，不再有新的数据产生，逐渐出现了信息冗余。例如，在访谈中，研究者不断重复听到同样的叙述，以此判断已达到数据饱和，因而停止数据搜集(Grady, 1998; Jackson et al. 2015)。数据饱和关注是否还会有新的信息出现，是在数据搜集阶段对资料饱和的判断。数据搜集处于整个研究过程的相对早期阶段，研究者基于“在随后的资料分析阶段，某些主题可能会出现”这一预设做出判断。但这种判断往往并不准确，因为在数据分析过程中，各种编码、编码属性及编码之间的关系还会不断变化，包括资料分析过程中编码的合并、意义扩充、意义排除等(Saunders et al., 2018)。所以，以数据饱和作为资料饱和的判断标准存在较为明显的弊端。

2.3 编码或主题饱和(code or thematic saturation)

编码或主题饱和有时也被称为类属饱和(categories saturation)，这一概念表示在数据分析过程中，编码或主题范围已基本确定，从新搜集的数据中只能发掘出与已有编码或主题相重复的内容，不再有新的编码或主题出现，编码书(codebook)已相对稳定(Urquhart, 2012; Hennink et al., 2017)。编码或主题饱和关注在数据分析阶段所产生的编码或主题在数量上的重复性，其判断标准为在数据分析过程中再无新的编码或主题出现。

编码或主题饱和与数据饱和相比可靠性有所进步。编码已经进入对数据的分析，而主题则是编码达到一定水平之后出现的结果。相比数据饱和仅仅根据数据或资料本身的重复性做

出判断，编码或主题饱和已经将一定程度的分析纳入其中。但以编码或主题重复作为资料饱和的判断依据仍然有问题，也容易出现虚假饱和，导致所获取的信息或资料不足以建构起完善的理论。这是因为编码或主题饱和仅仅提供了研究问题的大致轮廓，往往通过较少的访谈就能达到(Saunders et al., 2018)。因此，只达到编码或主题饱和还不够，仍需要增加更多数据，以便研究者充分理解问题的深度、丰富性和复杂性(Emmel, 2015; Hennink et al., 2017)。

与数据饱和一样，编码或主题饱和仅仅评估了编码或主题在数量或出现频度上的饱和，即只是基于所搜集材料的广度和范围进行饱和度判断，缺少对于编码或主题的意义评估。当某一主题首次从资料中浮现，研究者往往并不一定能达到对该主题的深入理解，还需要进一步搜集与分析数据，以发掘对于主题或概念类属更为丰富和深刻的认识(Hennink et al., 2017; Kerr et al., 2010)。

2.4 意义饱和(meaning saturation)

意义饱和是指在数据搜集和分析过程中，研究者已充分理解了所发展出的一系列编码或主题，关于编码或主题的含义及其之间的关系不再出现新的信息(Hennink et al., 2017)。Hennink 等(2017)认为，要达到意义饱和需要一个循环演进的抽样、搜集资料、分析资料的过程，持续监测数据的多样性、清晰度和深度，强调针对目前尚缺乏理解的信息、编码或主题进行数据搜集。意义饱和是在数据搜集和分析过程中基于编码或主题意义的完整性进行的资料饱和判断，着重于对资料的深度理解。使用这一指标判断饱和度，有利于发掘和呈现资料的完整意义。

Hennink 等(2017)比较了编码饱和与意义饱和，结果发现，在总共进行的 25 人次的深度访谈中，在第 9 次访谈达到了编码饱和，亦即将所有编码识别出来，确定了主要命题(thematic issues)的范围。然而，需要 16~24 次访谈才能达到意义饱和，即研究者真正认识了编码的含义，对研究问题有了丰富理解。Hennink 等认为，编码饱和仅代表着“听到过(heard it all)”，而达到意义饱和需要“理解了(understand it all)”。这再次说明，在进行饱和度检验时，仅仅依靠“有些编码或主题开始重复出现，不能继续发现其他新的编码或主题(Kerr et al., 2010)”作为判断标准是不够的，还应该继续判定每一个编码或主题的定义与内容是否得到了充足的发掘与认识，即“对每一个编码或主题的解释维度或理解开始重复出现，不再发现对编码或主题的其他解释，编码或主题因此达到了意义饱和。”

以上四种形式的资料饱和，关注了质性研究资料搜集和分析过程中的信息冗余或理论内涵(Sim et al., 2018)，它们分别出现在研究过程的不同阶段(Saunders et al., 2018)。数据饱和直

接关注资料搜集过程；编码或主题饱和则基于对数据的初步分析；意义饱和在数据搜集和分析(质性研究数据搜集与分析常常同时或滚动进行)的基础上，进一步注重资料分析的深度；而理论饱和则关注更高水平的理论建构，强调在理论建构过程中对概念类属及其属性(理论含义)发掘的充分性。

不同形式的资料饱和其判断的侧重点不同。数据饱和、编码或主题饱和基于数据、编码或主题的重复或出现频度进行判断；意义饱和基于编码或主题的意义深度进行判断；理论饱和基于新生理论中各概念属性的完整性、新生理论的自洽性及新生理论与以往理论的一致性进行判断。总的来说，数据饱和、编码或主题饱和着眼于所搜集资料的广度，意义饱和、理论饱和则着眼于研究资料的深度(Saunders et al., 2018)。意义饱和与理论饱和更接近，但二者又有本质区别。意义饱和关注的是编码或主题的意义深度；而理论饱和则进一步关注由编码或主题发展而来的新生理论的完整性、预测性与解释力。此外，理论饱和的概念源自并主要适用于扎根理论，而意义饱和则不限于扎根理论的语境，适用于对更多类型的质性研究资料饱和的判断(Hennink et al., 2017)。

3 资料饱和的判定方式及标准

3.1 理论饱和的判定

如前所述，理论饱和的概念来源于扎根理论。在扎根理论研究中，理论的发展与完善需要建立在系统地搜集和分析资料的基础上，是通过数据搜集与数据分析的相互作用实现的；扎根理论主要的分析方式是在资料与资料之间、理论与理论之间重复进行比较，基于资料和理论的相关性提炼出概念类属及其属性，又被称为“不断比较的方法” (陈向明, 2000)。不断比较是扎根理论研究的主要特征，理论饱和就依赖于不断进行的比较过程(Bowen, 2008)。在一系列比较完成之后，研究者勾勒出新的理论，并与早期搜集的资料、新搜集的资料进行持续比较，从而完善理论。当研究者发现理论可以解释大部分(或所有)原有资料或新资料时，即可判断该研究达到了理论饱和。

Bowen (2008) 以一项扎根理论研究为例，根据扎根理论研究的要求，进行了数据搜集、编码分析、理论抽样、持续比较等操作，并详细说明了对理论饱和的判定标准。他认为，满足以下四个条件方可认定研究资料达到理论饱和：1)相关概念类属反映在 70%以上的访谈中；2)受访者认同研究结果，对研究结果有较好的反馈；3)与前人研究结果相契合；4)新的受访者开始重复叙述与之前受访者相似的内容。Bowen 还建议，如果使用诸如 70%作为资料饱

和的判定标准，最好与其他形式的饱和标准结合使用。

3.2 数据饱和的判定

研究者根据资料搜集阶段访谈对象所叙述的内容以及对资料分析阶段可能出现的主题预设进行数据饱和度预判。当研究者重复听到同样的叙述，便开始考虑是否做出研究达到了数据饱和的判断(Jackson et al. 2015)。如前所述，在资料分析过程中，研究者的早期理解可能发生巨大变化，仅仅基于上述特征就判断研究达到资料饱和是不可靠的。但将数据饱和应用于对单次访谈的资料饱和评估，或对个体生命历程进行分析 (如个体心理传记等) 之类的质性研究中似乎是可行的。

3.3 编码或主题饱和的判定

编码或主题饱和是质性研究使用较多的资料饱和形式，对与此相关的判定标准的研究和讨论也最多。

3.3.1 以实证研究结果作为判定依据

有研究采用回顾性实证分析的方式，系统呈现了资料饱和度的评估、报告和验证过程，总结了研究达到编码或主题饱和所需的样本量。如 Guest 等 (2006)总共进行了 60 次深度访谈，在访谈结束后，以每 6 次访谈为一组，依据主题和重要主题的发展程度(依据主题所包含的编码数量确定重要主题)判断饱和度。结果发现，采用主题分析，前 6 次访谈结束后，主题的基本要素就已经呈现。在第 12 次访谈之后确定了 88%的主题与 97%的重要主题，编码书的结构趋于稳定，很少再需要更改，研究达到了主题饱和。Guest 等(2017)采用类似方法分析了 40 个焦点小组访谈获取的资料，结果发现，经过 3 个焦点小组访谈即可确定 84%的主题，6 个焦点团体访谈之后确定了 90%的主题。

Francis 等(2010) 探究了在基于理论的访谈研究中概念性类属(conceptual categories)的饱和，结果发现在经历了 17 人次的深度访谈之后，研究整体上达到了类属饱和。Coenen 等 (2012) 在研究中采用了最大变异抽样，验证焦点团体访谈中达到饱和所需要的样本量，结果发现经过 5 个焦点小组访谈，研究达到了编码饱和。Hennink 等 (2017, 2019)分别对半结构化深度访谈与焦点团体研究中所需的样本量进行了回顾性分析，结果发现，在总共进行的 25 次深度访谈中，第 9 次访谈达到了编码饱和；而在进行的 10 个焦点团体访谈中，第 4 次访谈就已达到编码饱和。类似的还有，研究者发现，在相对具有同质性的群体中，通过 16 个样本即可获得足够的共同主题，在跨文化背景下则需要 20~40 个样本(Hagaman & Wutich, 2017)；在对较抽象概念的探索中，通过对 12 个样本的访谈即可概括出 92%的编码(Ando et al.,

2014)。

上述研究大多采用后期回顾的分析方式，以研究获取的全部编码或主题为基数，通过计算某次访谈之后所获取的编码或主题在总数中的占比来判断资料饱和度，旨在基于多次实证研究的结果及研究者长期积累的经验，为后续研究提供标准参照。但是，在实际操作中，对于饱和度的判断是过程性的，即需要在资料搜集与分析的循环过程中评估饱和度，以便确定在何时停止抽样。回顾性分析仅相当于“事后聪明”或“马后炮”。尽管此类研究得出了众多关于样本量的实证经验，但正如很多研究者强调的，不能将其研究结果作为一般化取样要求直接应用于其他质性研究。因为资料饱和受到研究问题、研究目的、样本特征、取样均质性、访谈方式、编码特征、研究者以及数据特征等众多因素的影响，应根据研究的方法论特征、认识论立场以及研究资源等因素慎重确定样本量(Guest et al., 2006, Hennink et al., 2017)。

3.3.2 以是否还有新的编码或主题出现作为判定依据

有研究者倾向于追求数量化的指标，通过对比新获得的信息量与已获得的信息量判断资料饱和，即当某次访谈中新出现的编码数占已确定的编码总数的比例达到一定标准，即可判断研究达到资料饱和。Guest 等(2020)发现，在质性研究的数据搜集过程中，大部分新信息出现在早期，短期内通常遵循渐进曲线，在进行了一定量的数据搜集或分析之后，出现的新信息数量急剧下降，因此他们选择前 4 次(或 5 次、6 次)搜集到的基础数据集作为分母，以之后的每 2 次(或 3 次)访谈所获得的新信息量为分子，以 5%或 0%作为阈限值，判断资料饱和，即当新进行的 2 次访谈所获得的新信息数量占前 4 次访谈所获取的信息总量的比例小于 5%，即可判断研究达到了编码或主题饱和。其中，对饱和阈限值(5%或 0%)的选择参照了量化研究的显著性指标 p 值的设定。Guest 等提出，上述具体标准的选择是自主的，质性研究者可以根据研究现实自行确定。

通过计算新出现信息与已获信息之间的比率来判断资料饱和比较容易操作。研究者还可以借助于图表的辅助，根据研究推进过程中新编码或主题的出现做出符合研究目的的判断(Guest et al., 2020; Hennink et al., 2017, 2019)。例如 Hennink 等人使用条形图表示编码或主题在广度上的饱和，以新出现的编码数量为纵坐标，以访谈顺序为横坐标，条形图的高度表示在历次访谈中获得的新的编码或主题的数量(Hennink et al., 2017)。图 1 展现了该研究随着访谈的持续，编码或主题的发展过程。从中可以发现，首次访谈获取了大量编码，随后的访谈明显呈现收益递减的特征。从第 16 人次的访谈开始，连续出现不能提供任何新信息的情况，这表示编码或主题的范围已经基本确定，可以做出研究达到编码或主题饱和的判断。如果继

续进行访谈，虽然并不完全排除会有新信息出现的可能性，但质性研究并不追求对研究资料穷尽式的获取，只需获取相对足够的资料达到对概念与理论的发展和验证即可。

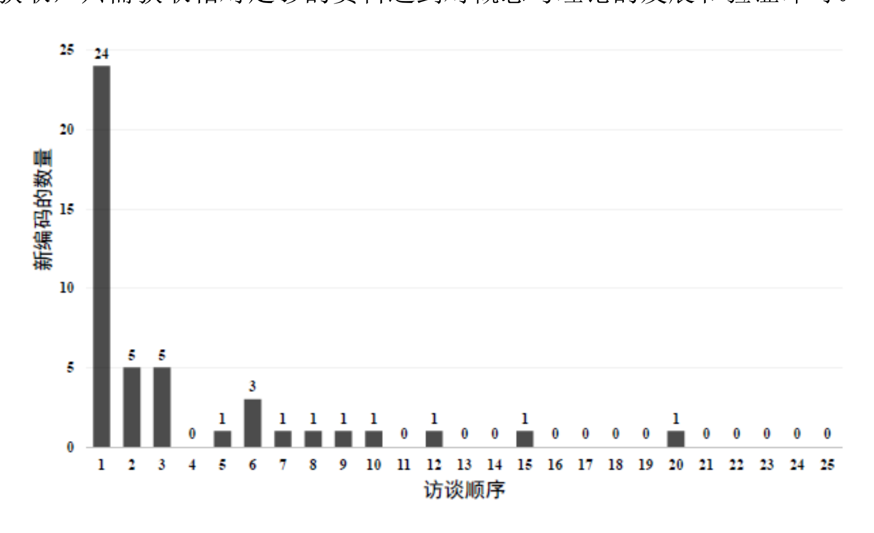


图 1 编码或主题发展过程图(Hennink et al., 2017)

Constantinou 等(2017)采用了比较方法(Comparative Method for Themes Saturation, CoMeTS) 判断主题饱和。首先，他们将对每一轮新的访谈资料的分析结果与之前的资料分析结果进行比较，明确最近一次访谈所获得的新主题及重复主题数量，当不再有新的主题出现时即视为达到资料饱和。然后，研究者将访谈资料的顺序打乱重新排序，再次进行资料分析，对主题饱和进行检验。结果发现，在按照访谈的自然顺序分析资料时，第 5 次访谈达到了主题饱和，而对访谈资料重新进行多次随机排序后，分别在第 7 次或第 8 次访谈才达到饱和。研究结果表明，访谈资料的分析顺序会影响对资料饱和的判断。应该指出，对访谈资料的收集与分析通常是滚动进行的，将访谈资料的顺序打乱重新排序并不符合质性研究的逻辑，但在判定过程中加入比较方法确实有助于对资料饱和做进一步确认。

3.3.3 以饱和度系数作为判定依据

有研究者试图用更复杂的计算公式探究质性研究的资料饱和度问题。Viet-Thi 等(2017)在开放式调查中使用蒙特卡罗模拟(Monte Carlo simulation)，又称随机模拟法，预测新加入的参与者所能提供的主题数量。Lowe 等(2018)基于数学统计模型测量质性研究的主题饱和，该模型提供了对数据集之间、数据集内部以及研究项目过程中饱和度的测量方式。Namey 等(2016) 演示了使用 bootstrap 模拟从每个质性研究数据集中随机生成 10000 个样本，计算达到主题饱和所需样本量的方法。另有研究者基于二项式分布(binomial distribution)提出样本量评估方式(Fugard & Potts, 2015; Galvin, 2015)。

以上多种资料饱和度的计算方式都依赖于概率理论与随机抽样(Fugard & Potts, 2015;

Galvin, 2015; Lowe, et al., 2018)。然而，由于质性研究采取非概率抽样方式(Guest, et al., 2020)，并且具有开放性等特征，使得它并不适用于概率论或者统计推断(Blaikie, 2018; Sim et al., 2018)。另外，受研究传统和知识范围的限制，复杂的数学统计分析模型既不方便人文社会科学研究者的操作和使用，也不利于相应的读者群接受和理解。

有研究者尝试引进 Jaccard 系数作为资料饱和度的判断指标(刘甜芳, 杨莉萍, 2018)。Jaccard 系数原本是指在给定的两个数据集 A 和 B 中，二者的交集与并集的比值，比值越大代表两个集合的相似度越高(张猛, 李玲娟, 2018)。Jaccard 系数的计算公式如下：

$$J(A,B)= |A\cap B|/|A\cup B|=|A\cap B|/(|A|+|B|-|A\cap B|)$$

在资料分析过程中，使用 Jaccard 系数可以计算不同轮次访谈所获取资料的编码相似性，新获取的数据集与之前已搜集的数据集相似度越大，Jaccard 系数越高，则表示资料编码的饱和度越高。这样的操作方式符合编码饱和的概念。在对英文资料的分析中，主要以英文单词作为分析单元计算 Jaccard 系数，而中文则以单个的汉字或编码作为识别单元。目前在操作过程中，Jaccard 系数大多只为一个报告值，至于其指标达到多少可以认定为编码饱和，并无统一标准，还有待进一步探索。此外，以 Jaccard 系数作为判断标准还存在其他问题。鉴于质性研究过程的持续性和探索性，以新获取的数据集 B 与此前已获取的所有数据的集合 A 合并作为分母，会使得分母偏大，造成 Jaccard 系数过小，从而对研究过程中资料饱和度的变化失去敏感性。

为了弥补 Jaccard 系数的不足，研究者构造出另一种操作简单、对资料饱和变化更为敏感的饱和度指标，称为 S(Saturation)系数，用以判断质性研究的编码饱和度。在给定的两个集合 A 和 B 中，B 代表新获得的数据集，A 代表在 B 之前已获取的所有数据的集合，两组数据编码的交集与数据集 B 的比值即为 S 系数。比值越大代表数据集 B 中的编码与之前集合的重复性越高。S 系数的计算公式如下：

$$S(A,B)=|A\cap B|/|B|$$

资料饱和度作为一种过程性评价指标，会随着研究过程的进展不断增加或累积。S 系数能很好地体现随着样本量增加，资料饱和度逐渐上升的趋势。在研究过程中每增加一次抽样，都可以重新计算资料饱和度，从而监测新增样本对于编码或主题发展的贡献程度。在对同质性样本的访谈中，深度访谈连续 3 次(焦点团体连续 2 次) S 系数达到 95%及以上可认定为达到编码饱和。这里 95%参照了量化研究的显著性指标 p 值的设定，而要求“连续 3 次”或“连续 2 次”，则是遵循了“适当追加抽样”的原则。但目前这依然只是经验性判断，有待

获得更多研究的验证与支持。

3.4 意义饱和的判定

Hennink 等 (2017, 2019) 采用回顾性实证分析的方式探究了深度访谈与焦点团体研究中达到意义饱和所需样本量，并与编码饱和做了对比。结果发现，需要 16~24 次个体访谈才能达到意义饱和，在第 10 次焦点团体访谈后勉强达到了意义饱和。通常情况下，达到意义饱和所需要的样本量远高于编码饱和。

研究者在资料搜集与分析的循环过程中借助于表格记录意义单元(二级编码)含义的发展变化，称为意义饱和网格 (Brod et al., 2009; Hennink et al., 2017, 2019)。如在表 1 中，第一列为发展出的意义单元，之后各列分别为第 1~6 次、7~9 次、10~12 次及第 12 次之后各阶段资料搜集为意义单元(或不同类属)所增加的新的内涵。

表 1 意义饱和网格(Hennink et al., 2017)

	意义单元含义			
意义单元	访谈 1~6	访谈 7~9	访谈 10~12	访谈 12 之后
感觉不错	没有疾病 (1)			
	感觉不错 (3)			
	病毒载量稳定(3)			
	有疾病所以就诊 (3)			
	有药物供应 (4)			
时间	出行/停留的时间 (1)	请假 (9)		
	会诊的时间 (1)			
	做其他事情的时间 (1)			
	诊所的时间安排 (1)			
	进行其他治疗的时间 (2)			
	等待时间 (3)			
	实验室测试时间 (3)			
	浪费时间 (3)			
工作承诺	长时间工作 (5)			雇主质疑请假(13)
	工作更重要 (6)			
	要在就诊时间工作 (6)			
	工作时间安排不灵活 (6)			
	工作请假规章制度 (6)			
	避免工作中的麻烦 (6)			
公开	污名公开 (1)	临床保密 (7)	工作场合公开 (11)	公开的负面经历 (13)
	选择性公开 (1)	向其他患者公开 (8)		公开曝光 (17)
	公开亲属与伙伴 (1)	向朋友公开 (9)		
	因污名隐藏 HIV (1)			
	害怕公开会带来暴力 (3)			
	推迟公开 (3)			
	其他公开情形 (6)			

意义饱和网格可以在资料搜集与分析的循环过程中建立。将发展出的意义单元逐一填入第一列，将新发现的意义单元的下级编码或自由节点按照各自出现的顺序分别填写在网格中的相应位置。如果在访谈中没有新的编码出现，则空出相应单元格。单个意义单元达到饱和的标志是，在后续进行的访谈中不再出现该意义单元新的下级编码或自由节点。当所有(或达到一定比例)的单元格都呈现为空白，代表随着访谈继续下去不会再有新的信息增加，即可判断研究资料达到了意义饱和。在表 1 中，第 2 个意义单元“时间”的属性到第 9 次访谈已基本确定，之后没有继续增加，说明该意义单元已达到饱和。需要提醒的是，在质性研究过程中，编码始终处于不断调整状态，这一点在意义饱和网格中能够体现出来。

意义饱和网格记录了意义单元及其含义的发展过程，可以帮助研究者判断特定的意义单元是否已经达到饱和。但这类网格大多只能保存在研究者的研究备忘中，由于内容太多，过于庞杂，在期刊论文中很难完整呈现。Hennink 等(2017)因此创造了另一种更具参考价值的意义饱和呈现方式，见图 2。其中，横坐标为访谈顺序，纵坐标分别呈现不同的意义单元，空心圆表示某个意义单元被正式创建的位置，实心圆表示某意义单元达到意义饱和的位置，即在后续访谈中有关该意义单元的新信息不再出现，这时可认定该意义单元的相关信息已达到饱和。

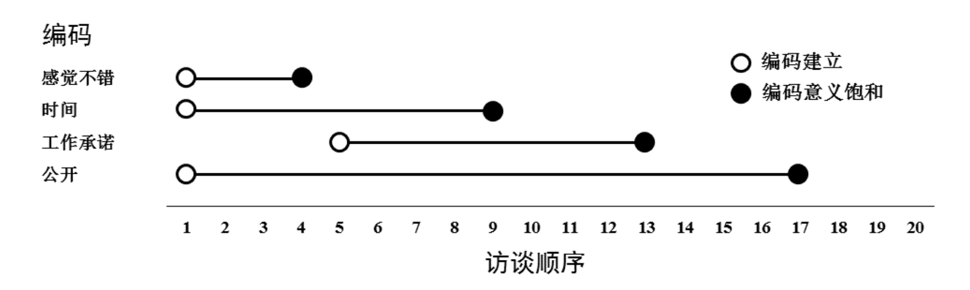


图 2　意义饱和坐标图(Hennink et al., 2017)

4 讨论

在质性研究中强调资料饱和具有三方面意义。其一，评估研究资料的充足性，以指导抽样过程。其二，帮助研究者判断对研究问题及相关概念的认识是否深入，是研究质量的体现。其三，帮助研究者在论文中报告研究取样的充分性，接受读者的审查，也便于后续研究者进一步跟进研究。但在实际操作过程中，不同形式的资料饱和常常被混淆，不恰当的评估方式或模糊操作等问题普遍存在。

4.1 质性研究达到资料饱和的样本量标准不应统一设定，其检验需嵌入具体研究过程

对于资料饱和的判定，某些实证研究的结果或统计学方法试图在研究之前就给出达到资料饱和的样本量标准。但是，如前所述，质性研究的资料饱和受诸多因素的影响，其中每项研究都有其个性或特殊性。例如，在访谈研究中，主题与受访者紧密相连，受访者的身份和所处的语境帮助赋予主题意义和重要性，主题是受访者的“属性”，二者相互对应(Byrne, 2015; Hammersley, 2015; Sim et al., 2018)。而在资料分析过程中，主题也不是一成不变的，主题的命名、内涵、外延及其对研究问题的理论贡献都在不断发展，特定主题的重要性及其与研究问题的相关度也随之改变(Hammersley, 2015)。因此，对研究问题探索的深度与样本量之间并非服从简单的线性关系，受访者人数不能单独作为判断资料饱和的依据(Sim et al., 2018)，达到研究资料饱和的样本量也不能在研究开始之前确定(Braun & Clarke, 2021)。

在具体研究过程中，研究者应基于本研究的特点，综合考虑多种资料饱和形式。如使用编码或主题饱和评估研究资料的广度和范围，使用意义饱和评估编码或主题被探索的深度。如果该研究以建构理论为目的，还应该通过持续性比较对理论饱和加以检验。将检验嵌入研究过程，综合使用多种形式的资料饱和，在国外也有一些研究案例，如数据饱和与理论饱和结合使用(Goulding, 2005; Morse, 2015)。

4.2 质性研究的资料饱和存在逻辑上的不确定性，适当追加抽样有助于进一步确认

资料饱和是基于已获取的数据对未来访谈数据获取情况所做出的预测。这里“逻辑上的不确定性”是指，根据当前资料搜集和分析的情况预测继续进行数据搜集和分析的必要性，这种预测在一定程度上依赖于研究者的主观判断，其准确性并不能得到进一步的证明(Saunders et al., 2018)。尽管研究者可以清晰地记录和呈现资料搜集与分析过程，为研究的资料饱和提供证据，但质性研究的资料饱和很难说是一个绝对准确的判断。

正因为资料饱和在逻辑上存在不确定性，有研究者认为，应该在资料达到或基本达到饱和之后再适当追加抽样，以对资料饱和加以验证。例如，Jassim 和 Whitford(2014)在进行了 10 次访谈之后，发现研究资料已经达到了主题饱和，但仍继续追加了 2 次访谈，以对资料饱和进行确认。类似的方式也被其他研究者所采用(Bragaru et al. 2013; Jackson et al., 2000; Vandecasteele et al., 2015)。但 Saunders 等(2018)指出，过度抽样的做法在一定程度上也存在问题，可能会造成资料饱和概念的模糊性。尽管如此，适当追加抽样仍不失为应对资料饱和逻辑不确定性的有效策略。通常在做出研究资料饱和或基本饱和的判断之后，研究者可以根据需要再继续针对同质性样本追加 2~3 次个体访谈或 1~2 次焦点小组访谈，对资料饱和做

进一步的确认。

4.3 资料饱和作为考察研究质量的一项重要指标，并不适用于所有质性研究

质性研究是一个“伞概念 (umbrella concept)”，包含各种不同的研究方法，如扎根理论、现象学研究、话语分析、主题分析、民族志、叙事研究、焦点团体、生命史与心理传记等 (何吴明, 郑剑虹, 2019)。有研究者认为，资料饱和的概念并不适用于叙事分析、解释现象学分析等(Marshall & Long, 2010; O’Reilly & Parker, 2013; van Manen et al. 2016)。主题分析、扎根理论、焦点团体等采用归纳思维，从众多受访者那里搜集信息，关注理论发展的充分性，因此比较符合资料饱和的操作性定义；而心理传记、叙事分析等则聚焦于个体，更加关注个体故事的完整性，现有的资料饱和概念似乎并不包含此类含义(Saunders et al., 2018)。至于解释现象学分析，虽然也会采用提炼主题、明确主题间关系等方式进行资料分析，但更重视获取完整、丰富的个体关于生活经验的理解，具有特则取向，强调对每个案例进行详细的审视(侯力琪 等, 2019)，这类研究是否适用于资料饱和的概念，还需要进一步探索和商榷。因此，有研究者担心，如果将资料饱和这一概念无差别地应用到所有质性研究中，可能导致失去其一致性效用(Saunders et al., 2018)。未来需要进一步关注对不同类型质性研究的资料充足性的考察、判断与检验。

5 结语

质性研究并非是在传统实证研究的框架内新增的一种研究方法，而是代表了心理学中的一种全新方法论。它在研究选题、文献综述、研究关系的建立、资料搜集、资料分析、结果与讨论、效度检验、研究报告等各个环节，都表现出与实证研究不同的特点。资料饱和及其检验只是其中一个研究环节。只有针对质性研究过程中的每一环节开展深入细致的研究，才能提高质性研究操作的规范性，从而提高研究质量。

参考文献

陈向明. (2000). *质的研究方法与社会科学研究*. 北京: 教育科学出版社.

侯力琪, 唐信峰, 何丽, 贾晓明. (2019). 解释现象学分析在中国的运用：系统评价及指南. *心理科学进展, 27*(11), 1826–1841.

何吴明, 郑剑虹. (2019). 心理学质性研究：历史、现状和展望. *心理科学, 42*(04), 1017–1023.

刘甜芳, 杨莉萍. (2018). 基于年龄建构的老化心理——一项深度访谈研究报告. *心理科学, 41*(05),

1207–1213.

苏达比, R. (2006). 扎根理论之所非 (曾宪聚, 韩巍 译). 见 徐淑英, 任兵, 吕力 (编). (2016).*管理理论建构论文集*. (pp. 145–162). 北京: 北京大学出版社.

张猛, 李玲娟. (2018). 基于改进的 Jaccard 相似系数矩阵的社团划分算法. *南京邮电大学学报(自然科学版)*, *38*(06), 96–102.

Ando, H., Cousins, R., & Young, C. (2014). Achieving saturation in thematic analysis: Development and refinement of a codebook. *Comprehensive Psychology, 3*, 1–7.

Bowen, G. A. (2008). Naturalistic inquiry and the saturation concept: A research note. *Qualitative Research, 8*(1), 137–152.

Byrne, D. (2015). Response to Fugard and Potts: Supporting thinking on sample sizes for thematic analyses: A quantitative tool. *International Journal of Social Research Methodology, 18*(6), 689–691.

Blaikie, N. (2018). Confounding issues related to determining sample size in qualitative research. *International Journal of Social Research Methodology, 21*(5), 635–641.

Braun, V., & Clarke, V. (2021). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health, 13*(2), 201–216.

Brod, M., Tesler, L. E., & Christensen, T. L. (2009). Qualitative research and content validity: Developing best practices based on science and experience. *Quality of Life Research, 18*(9), 1263–1278.

Bragaru, M., van Wilgen, C. P., Geertzen, J. H., Ruijs, S. G., Dijkstra, P. U., & Dekker, R. (2013). Barriers and facilitators of participation in sports: A qualitative study on Dutch individuals with lower limb amputation. *PloS one, 8*(3), e59881.

Coenen, M., Coenen, T., Stamm, A., Stucki, G., & Cieza, A. (2012). Individual interviews and focus groups with patients with rheumatoid arthritis: A comparison of two qualitative methods. *Quality of Life Research, 21*(2), 359–370.

Constantinou, C. S., Georgiou, M., & Perdikogianni, M. (2017). A comparative method for themes saturation (CoMeTS) in qualitative interviews. *Qualitative research, 17*(5), 571–588.

Curtis, S., Gesler, W., Smith, G., & Washburn, S. (2000). Approaches to sampling and case selection in qualitative research: Examples in the geography of health. *Social Science and Medicine, 50*(7-8), 1001–1014.

Carlsen, B., & Glenton, C. (2011). What about N? A methodological study of sample-size reporting in focus group studies. *BMC Medical Research Methodology, 11*(1), 1–10.

Corbin, J., & Strauss, A. (2014). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Thousand Oaks, CA: Sage Publications, Inc.

Emmel, N. (2015). Themes, variables, and the limits to calculating sample size in qualitative research: A response to Fugard and Potts. *International Journal of Social Research Methodology, 18*(6), 685–686.

Francis, J., Johnson, M., Robertson, C., Glidewell, L., Entwistle,V., Eccles, M., & Grimshaw, J. (2010). What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology and Health, 25*(10), 1229–1245.

Fusch, P. I., & Ness, L. R. (2015). Are we there yet? Data saturation in qualitative research. *The Qualitative Report, 20*(9), 1408–1416.

Fugard, A. J., & Potts, H. W. (2015). Supporting thinking on sample sizes for thematic analyses: A quantitative tool. *International Journal of Social Research Methodology, 18*(6), 669–684.

Grady, M. P. (1998). *Qualitative and action research: A practitioner handbook.* Bloomington, IN: Phi Delta Kappa Educational Foundation.

Goulding, C. (2005). Grounded theory, ethnography and phenomenology A comparative analysis of three qualitative strategies for marketing research. *European Journal of Marketing, 39*(3/4), 294–308.

Galvin, R. (2015). How many interviews are enough? Do qualitative interviews in building energy consumption research produce reliable knowledge? *Journal of Building Engineering, 1*(1), 2–12.

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods, 18*(1), 59–82.

Guest, G., Namey, E., & Chen, M. (2020). A simple method to assess and report thematic saturation in qualitative research. *PloS One, 15*(5). e0232076.

Guest, G., Namey, E., & McKenna, K. (2017). How many focus groups are enough? Building an evidence base for nonprobability sample sizes. *Field Methods, 29*(1), 3–22.

Glaser, B. G., Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine Transaction.

Hammersley, M. (2015). Sampling and thematic analysis: A response to Fugard and Potts. *International Journal of Social Research Methodology, 18*(6), 687–688.

Hennink, M. M., Kaiser, B. N., & Marconi, V. C. (2017). Code saturation versus meaning saturation: How many interviews are enough? *Qualitative Health Research, 27*(4), 591–608.

Hennink, M. M., Kaiser, B. N., & Weber, M. B. (2019). What influences saturation? Estimating sample sizes in

focus group research. *Qualitative Health Research, 29*(10), 1483–1496.

Hagaman, A. K., & Wutich, A. (2017). How many interviews are enough to identify metathemes in multisited and cross-cultural research? Another perspective on Guest, Bunce, and Johnson's (2006) Landmark Study. *Field Methods, 29*(1), 23–41.

Jackson, D., Daly, J., Davidson, P., Elliott, D., Cameron-Traub, E., Wade, V., Chin, C., & Salamonson, Y. (2000). Women recovering from first-time myocardial infarction (MI): A feminist qualitative study. *Journal of Advanced Nursing, 32*(6), 1403–1411.

Jackson, M., Harrison, P., Swinburn, B., & Lawrence, M. (2015). Using a qualitative vignette to explore a complex public health issue. *Qualitative Health Research, 25*(10), 1395–1409.

Jassim, G. A., & Whitford, D. L. (2014). Understanding the experiences and quality of life issues of Bahraini women with breast cancer. *Social Science and Medicine, 107,* 189–195.

Kerr, C., Nixon, A., & Wild, D. (2010). Assessing and demonstrating data saturation in qualitative inquiry supporting patient-reported outcomes research. *Expert Review of Pharmacoeconomics and Outcomes Research, 10*(3), 269–281.

Levitt, H. M., Bamberg, M., Creswell, J. W., Frost, D. M., Josselson, R., & Suárez-Orozco, C. (2018). Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*(1), 26-46.

Lowe, A., Norris, A. C., Farris, A. J., & Babbage, D. R. (2018). Quantifying thematic saturation in qualitative data analysis. *Field Methods, 30*(3), 191–207.

Morse, J. M. (2015). "Data were saturated . . . ". *Qualitative Health Research, 25*(5), 587–588.

Marshall, B., Cardon, P., Poddar, A., & Fontenot, R. (2013). Does sample size matter in qualitative research? A review of qualitative interviews in IS research. *Journal of Computer Information Systems, 54*(1), 11–22.

Marshall, V., & Long, B. C. (2010). Coping processes as revealed in the stories of mothers of children with autism. *Qualitative Health Research, 20*(1), 105–116.

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research, 26*(13), 1753–1760.

Namey, E., Guest, G., McKenna, K., & Chen, M. (2016). Evaluating bang for the buck: A cost-effectiveness comparison between individual interviews and focus groups based on thematic saturation levels. *American Journal of Evaluation, 37*(3), 425–440.

O'Reilly, M., & Parker, N. (2013). 'Unsatisfactory Saturation': A critical exploration of the notion of saturated sample sizes in qualitative research. *Qualitative Research, 13*(2), 190–197.

Sim, J., Saunders, B., Waterfield, J. & Kingstone, T. (2018) Can sample size in qualitative research be determined a priori? *International Journal of Social Research Methodology, 21*(5), 619–634.

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2018). Saturation in qualitative research: Exploring its conceptualization and operationalization. *Quality and Quantity, 52*(4), 1893–1907.

van Manen, M., Higgins, I., & van der Riet, P. (2016). A conversation with Max van Manen on phenomenology in its original sense. *Nursing and Health Sciences, 18*(1), 4–7.

Vandecasteele, T., Debyser, B., Van Hecke, A., De Backer, T., Beeckman, D., & Verhaeghe, S. (2015). Nurses' perceptions of transgressive behaviour in care relationships: A qualitative study. *Journal of Advanced Nursing, 71*(12), 2786–2798.

Viet-Thi, T., Porcher, R., Viet-Chi, T., & Ravaud, P. (2017). Predicting data saturation in qualitative surveys with mathematical models from ecological research. *Journal of Clinical Epidemiology, 82,* 71–78.

Urquhart, C. (2012). *Grounded theory for qualitative research: A practical guide*. Thousand Oaks, CA: Sage Publications, Inc.

Concepts and evaluation of saturation in qualitative research

YANG Liping, QI Lidong, ZHANG Bo

(*School of Psychology, Nanjing Normal University, Nanjing 210024, China*)

Abstract: In qualitative research, saturation is usually used to assess the adequacy of research data. However, in research practice, there are various forms of saturation, and the relationships between them are complicated and ambiguous. Previous studies lack operational description and practical guidance for the evaluation of saturation and its reports. Considering its importance to qualitative research, this study clarifies the concepts of saturation. As a sign that the sample size of a qualitative study has met the research needs, based on the time sequence of their occurrence during the research process, saturation can be distinguished into four types: data saturation, code or

thematic saturation, meaning saturation, and theoretical saturation. Each of them has its specific connotations, evaluation methods and judgment criteria. Some problems are discussed in this study. 1) The sample size to reach saturation should not be set uniformly, and it must be embedded in the specific research process; 2) Due to the logical uncertainty of saturation, a little oversampling would be helpful; 3) As an important index to evaluate the quality of a qualitative research, saturation test cannot be applicable to all forms of qualitative research.
Keywords: qualitative research, saturation, evaluation methods, judgment criteria